



BACHELORARBEIT

Herr
Christian Baldauf

**Implementierung eines
Algorithmus zur automatisierten
Lane-Entzerrung auf
Gelelektrophorese-Blots**

2014

BACHELORARBEIT

Implementierung eines Algorithmus zur automatisierten Lane-Entzerrung auf Gelelektrophorese-Blots

Autor:
Christian Baldauf

Studiengang:
Informatik

Seminargruppe:
If10w1-B

Erstprüfer:
Prof. Dr. Dirk Labudde

Zweitprüfer:
Dipl.-Ing. Heiko Mixtacki

Mittweida, 2014

Bibliografische Angaben

Baldauf, Christian: Implementierung eines Algorithmus zur automatisierten Lane-Entzerrung auf Gelelektrophorese-Blots, 29 Seiten, 11 Abbildungen, Hochschule Mittweida, University of Applied Sciences, Fakultät Mathematik/Naturwissenschaften/Informatik

Bachelorarbeit, 2014

Referat

Diese Bachelorarbeit beschäftigt sich mit Entwicklung und Implementierung eines Algorithmus zur automatisierten Entzerrung von Gelelektrophorese-Blots. Dazu werden zuerst Algorithmen für die Detektion der Lanes und Banden vorgestellt. Danach werden Möglichkeiten für die Erkennung von Verzerrungen, die Korrektur des Laneverlaufs und schließlich der Entzerrung der Lane vorgestellt.

I. Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	II
1 Einleitung	1
1.1 Motivation	1
1.2 Voraussetzungen	1
2 Grundlagen der Elektrophorese	3
2.1 Physikalische Grundlagen	4
2.1.1 Elektrophoretische Mobilität	4
2.1.2 Elektroendosmose	5
2.2 Verwendete Gele	5
2.3 Visualisierung	5
2.3.1 Anfärbung	5
2.3.2 Blotting	7
2.4 Verzerrungen	7
3 Lanedetektion	9
3.1 Bildvorverarbeitung	9
3.1.1 Rauschunterdrückung	9
3.1.2 Kontrasterhöhung	9
3.2 Detektion von Lanegrenzen	10
3.2.1 Finden von Segmenten der Lanegrenzen	10
3.2.2 Zusammensetzen der Lanegrenzen	12
3.2.3 Fehlerkorrektur in den Lanegrenzen	12
3.3 Bestimmung der Lanes	14
3.3.1 Finden von Lanekandidaten	14
3.3.2 Prüfung der gefundenen Lanekandidaten auf Richtigkeit	14
3.3.3 Suchen der Positionen von nicht erkannten Lanes	15
3.3.4 Erneutes Suchen nach Lanes	15
3.3.5 Abschneiden und Auffüllen der Lanes	16

4	Lane-Entzerrung	17
4.1	Feststellen der zu entzerrenden Lanes	17
4.2	Bandendetektion	17
4.2.1	Vorbereitung	17
4.2.2	Finden der Banden	18
4.3	Lane-Entzerrung	19
4.3.1	Verlauf der Entzerzten Lane	19
4.3.2	Auswahl der Passpunkte	19
4.3.3	Zerlegung der Lane	20
4.3.4	Transformation	21
4.3.5	Resampling	21
5	Überblick über die Implementierung	23
5.1	Aufruf	23
5.2	Übersicht über die Klassen	23
6	Auswertung	25
6.1	Zusammenfassung	25
6.2	Ausblick	25
	Literaturverzeichnis	27

II. Abbildungsverzeichnis

2.1 Gelelektrophorese (Ana98)	3
2.2 Elektroendosmose (Wes06)	5
2.3 Eingefärbte Gele	6
2.4 Gel mit stark verzerrter Lane	7
3.1 Bildvorverarbeitung	11
3.2 Segmente der Lanegrenzen	12
3.3 Lanegrenzen	13
3.4 Lanes	16
4.1 Vorbereitung der Entzerrung	19
4.2 Kubische Faltung	22
4.3 Entzerrtes Bild	22

1 Einleitung

1.1 Motivation

Die Gelelektrophorese ist ein sehr wichtiges Verfahren in der Molekularbiologie, der Biochemie, der Genetik und der Forensik. Es können verschieden Proteine oder DNA-Fragmente innerhalb eines Stoffgemisches voneinander getrennt, und somit die Zusammensetzung des Stoffgemisches analysiert werden.

Die entstandenen Trennmuster werden meist digitalisiert, um mit spezieller Software analysiert zu werden. Verzerrungen einzelner Banden oder ganzer Lanes können zu Schwierigkeiten bei der Auswertung führen, oder die Analyse komplett unmöglich machen. Ein Verfahren zur Korrektur solcher Verzerrungen kann dazu beitragen, solche Gele dennoch auswerten zu können und somit eine neue Messung überflüssig zu machen. Insbesondere wenn dieses Verfahren automatisiert abläuft, kann eine deutliche Zeitersparnis erreicht werden.

1.2 Voraussetzungen

Der in dieser Arbeit vorgestellte Algorithmus kann Gelbilder nur analysieren, wenn diese bestimmte Voraussetzungen erfüllen. So ist es notwendig, dass es sich um ein Graustufenbild beliebiger Farbtiefe handelt, bei dem Hintergrund dunkel ist, während sich die Banden hell darauf abheben. Des weiteren ist es notwendig, dass die Lanes senkrecht vom oberen Bildrand in Richtung des unteren Bildrandes verlaufen.

Allerdings können auch Gelbilder, die diese Voraussetzungen nicht erfüllen, entsprechend vorbereitet werden, dass sie doch von diesem Algorithmus analysiert werden können. So können Bilder mit waagerecht verlaufenden Lanes gedreht werden und Bilder mit dunklen Banden auf hellem Hintergrund können invertiert werden (siehe (Kum06)). Es können auch Farbbilder verwendet werden, wenn diese vorher in ein Graustufenbild konvertiert werden. Der Algorithmus kann sie dann auf die gleiche Weise bearbeiten, wie gewöhnliche Graustufenbilder. Lediglich die mögliche Entzerrung muss dann für die verschiedenen Farbkanäle getrennt durchgeführt werden.

2 Grundlagen der Elektrophorese

Als Elektrophorese versteht man einen Vorgang, bei dem verschiedene Moleküle innerhalb eines Gemisches mit Hilfe eines elektrischen Feldes getrennt werden. Da sich die Moleküle i. d. R. in einer wässrigen Lösung befinden, bewegen sich geladene Moleküle in die Richtung der Elektrode mit der entgegengesetzten Ladung. Aufgrund von unterschiedlichen Massen und Ladungen bewegen sich die Moleküle mit unterschiedlichen Geschwindigkeiten und können somit getrennt werden. (Wes06)

Es entsteht ein Trennmuster, bei dem die einzelnen Proben parallel zueinander als Lanes durch das Bild verlaufen. Die verschiedenen Komponenten des eingebrachten Stoffgemisches zeigen sich dabei in individuellen Bereichen, Banden genannt, innerhalb der Lanes.

Diese Arbeit befasst sich nur mit der Gelelektrophorese. Dabei werden Gele verwendet, die als Träger die elektrophoretische Trennung stabilisieren (Ric03). Die Gelplatten werden in vertikalen und horizontalen Systemen eingesetzt. Ein Überblick über die Vorteile und Nachteile beider Systeme ist bei Westermeier (Wes06) zu finden.

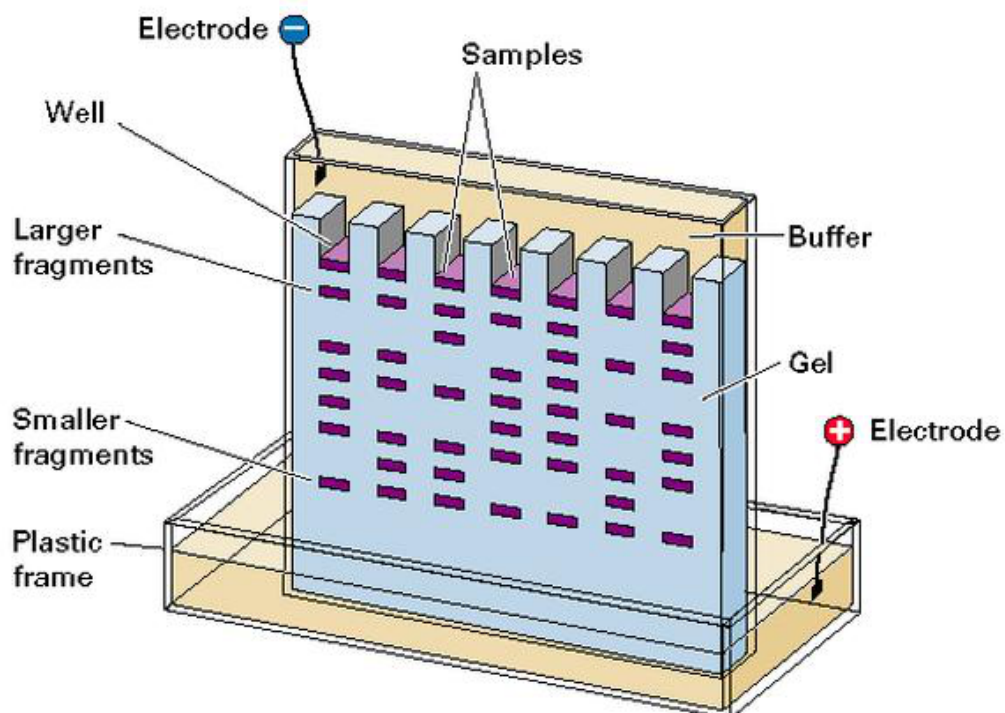


Abbildung 2.1: Gelelektrophorese (Ana98)

2.1 Physikalische Grundlagen

2.1.1 Elektrophoretische Mobilität

Die Geschwindigkeit v_i , mit der sich ein Ion i in einem elektrischen Feld bewegt, ist proportional zur Feldstärke E und damit der angelegten Spannung (Sch10):

$$v_i = \mu_i \cdot E$$

Der Proportionalitätsfaktor μ_i wird als elektrophoretische Mobilität des Ions i bezeichnet und stellt eine charakteristische Konstante des Ions bezogen auf den verwendeten Elektrolyt dar (Sch10).

Auf das Ion i wirkt zum einen die beschleunigende elektrische Kraft F_e und zum anderen die entgegengesetzt gerichtete Reibungskraft F_R . Dabei gilt (Sch10):

$$F_E = z_i \cdot e_0 \cdot E$$

wobei e_0 die Elementarladung und z_i die Ladungszahl des Ions i ist.

Die Reibungskraft F_R kann durch das Stoke'sche Gesetz angenähert werden, wenn das Ion vereinfachend kugelförmig betrachtet wird. Dadurch ergibt sich (Sch10):

$$F_R = 6 \cdot \pi \cdot \eta \cdot r_i \cdot v_i$$

dabei ist r_i der Radius des Ions, v_i seine Wanderungsgeschwindigkeit und η die dynamische Viskosität der das Ion umgebenden Flüssigkeit.

Im Kräftegleichgewicht ergibt sich damit (Sch10):

$$v_i = \frac{z_i \cdot e_0 \cdot E}{6 \cdot \pi \cdot \eta \cdot r_i}$$

Somit ergibt sich die elektrophoretische Mobilität zu (Sch10):

$$\mu_i = \frac{z_i \cdot e_0}{6 \cdot \pi \cdot \eta \cdot r_i}$$

„Die elektrophoretische Mobilität μ_i kann nicht berechnet werden, da der Stoke'sche Radius r_i bei einem gelösten Ion auch dessen Solvationshülle enthält und nicht mit dessen messbaren Radius im Kristallgitter korreliert werden kann. Auch die effektive Ladungszahl z_i wird durch den umgebenden Puffer beeinflusst.“ (Sch10)

2.1.2 Elektroendosmose

Die Gelmatrix kann geladene Gruppen enthalten, z. B. Carboxygruppen in Agarose. Diese Gruppen werden in basischen und neutralen Puffern ionisiert und im elektrischen Feld von der Anode angezogen. Da sie allerdings in der Matrix fixiert sind, können sie sich nicht bewegen und es entsteht eine Gegenbewegung von H_3O^+ Ionen in Richtung der Kathode. Als Resultat kann es zu verschwommenen Banden und dem Trocknen des Gels im Bereich um die Anode führen. (Wes06)

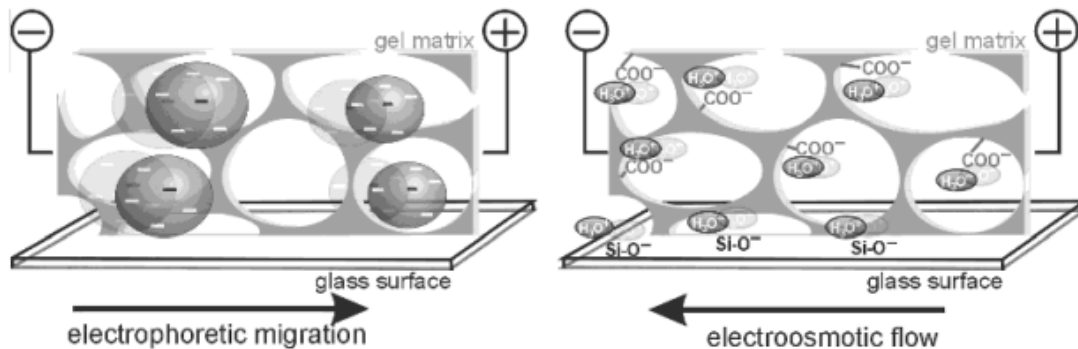


Abbildung 2.2: Elektroendosmose (Wes06)

2.2 Verwendete Gele

Heute werden hauptsächlich Gele aus Agarose oder Polyacrylamid verwendet, denn sie haben eine weitgehend chemisch inerte Natur, ein hohes Auflösungsvermögen für makromolekulare Analyten und eine durchsichtige Konsistenz. Weiterhin lässt sich der Vernetzungsgrad beider Polymere und somit die Porengröße im Gel variieren und so der Größenordnung der zu trennenden Analyten anpassen. Polyacrylamidgele werden dabei für die Analyse von Proteinen, DNA-Protein-Komplexen und freier DNA verwendet, während Agarosegele schwerpunktmäßig für die Analyse von Nukleinsäuren, insbesondere von DNA, verwendet werden. (Ric03)

2.3 Visualisierung

Da die meisten untersuchten Moleküle farblos sind, müssen sie visualisiert werden, bevor sie ausgewertet werden können.

2.3.1 Anfärbung

Proteine werden überwiegend mit Coomassie Brilliant-Blau angefärbt. Nach der Elektrophorese wird dabei das Gel mit den Proteinen in eine saure alkoholische Lösung des

Farbstoffs getaucht. Dadurch denaturieren die Proteine, werden im Gel fixiert, damit sie nicht auswaschen, und es wird dem Farbstoff ermöglicht, sich an die Proteine zu binden. Nachdem der überschüssige Farbstoff gewaschen wurde, sind die Proteine als blaue Banden sichtbar. Auf diese Weise können Proteine noch mit 0,1 - 1,0 µg im Gel visualisiert werden. (HH11)

Die Empfindlichkeit von Coomassie Brilliant-Blau kann man mit einer anderen Methode, der Silberfärbung, noch um das 100fache übertreffen. Zuerst werden die Proteine mit Hilfe von Essigsäure oder Trichloressigsäure auf dem Gel fixiert. Danach folgt eine Behandlung mit Glutaraldehyd oder Formamid und dann mit Silbernitratlösung. Dabei binden Silberionen an die Aminosäure-Reste und werden zu metallischem Silber reduziert und bilden Silberkerne. Durch ein zugegebenes starkes Reduktionsmittel werden die Silberionen im Gel in metallisches Silber überführt. Diese Umwandlung findet bevorzugt dort statt, wo sich bereits Silberkerne befinden. Dadurch färben sich die Proteinbanden schnell braun oder schwarz, methodenbedingt auch orange, rot oder gelb. Die überschüssigen Silberionen müssen ausgewaschen werden, bevor sie alle in metallisches Silber übergegangen sind, da sonst eine starke Hintergrundfärbung auftritt, die die Proteinbanden überlagern kann. (Ric03)

Zur Einfärbung von DNA und RNA wird hauptsächlich Ethidiumbromid verwendet. Ethidiumbromid kann bereits während der Elektrophorese im Gel präsent sein, allerdings erhält man bessere Ergebnisse, wenn das Gel erst nach der Elektrophorese mit dem Farbstoff versetzt wird. Das Ethidiumbromid lagert sich zwischen die Basen der DNA ein. Unter ultraviolettem Licht ist die eingefärbte DNA dann als rot-orange Banden sichtbar. Es können mit dieser Methode Banden detektiert werden, die 10 ng DNA enthalten. (KG05)

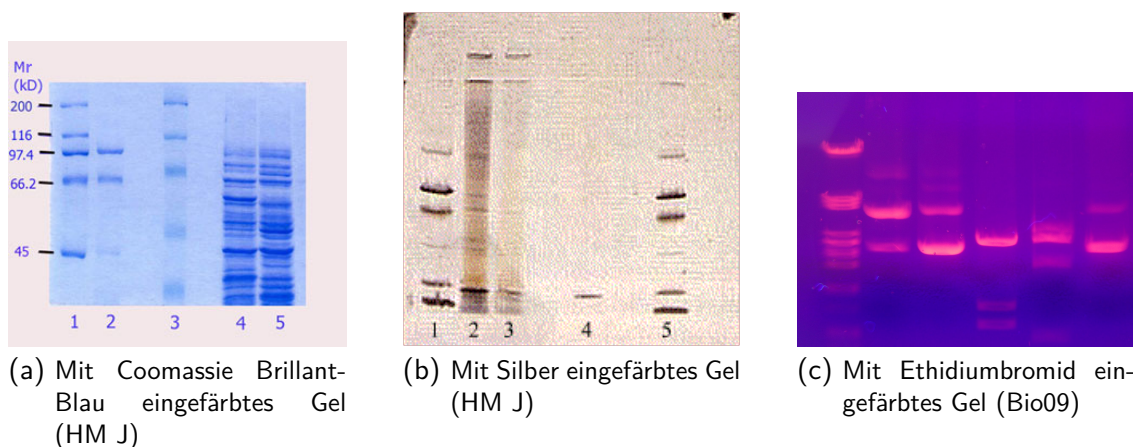


Abbildung 2.3: Eingefärbte Gele

2.3.2 Blotting

Zur weiteren Analyse kann es notwendig sein, die DNA, RNA oder Proteine unter Beibehaltung des Trennmusters aus der Gelmatrix auf eine undurchlässige, aber bindungsaktive Membran aus Nitrocellulose oder Nylon zu überführen. Eine Variante dafür ist das Kapillar-Blotting. Dabei wird das Gel auf feuchtes Filterpapier gelegt und darauf die Membran und trockenes Filterpapier. Die Moleküle werden mit dem Fluss des Puffers vom feuchten zum trockenen Filterpapier übertragen. Schließlich wird die Membran mit einer blockierenden Lösung getränkt, die hohe Konzentrationen von DNA, RNA oder Proteinen enthält. Dadurch wird verhindert, dass sich später die Sonde einfach mit der Membran verbindet (LTD⁺13). Zum Finden bestimmter Sequenzen kann dann Western Blotting für Proteine (HH11), Southern Blotting für DNA (Ric03) und Northern Blotting für RNA (Kni06) verwendet werden.

2.4 Verzerrungen

Durch die Verzerrung einzelner Banden oder ganzer Lanes wird die Auswertung dieser erschwert oder gänzlich unmöglich gemacht. Besonders wichtig hierbei sind die gleichmäßige Erwärmung des Gels und ein horizontaler Pufferspiegel. Unregelmäßigkeiten bei diesen Faktoren können zu Verzerrungen der Banden führen (WKC90). Die Verzerrung einer Lane kann auftreten, wenn Proben mit hohen Konzentrationen bestimmter Salze (z. B. NaCl) kontaminiert sind.

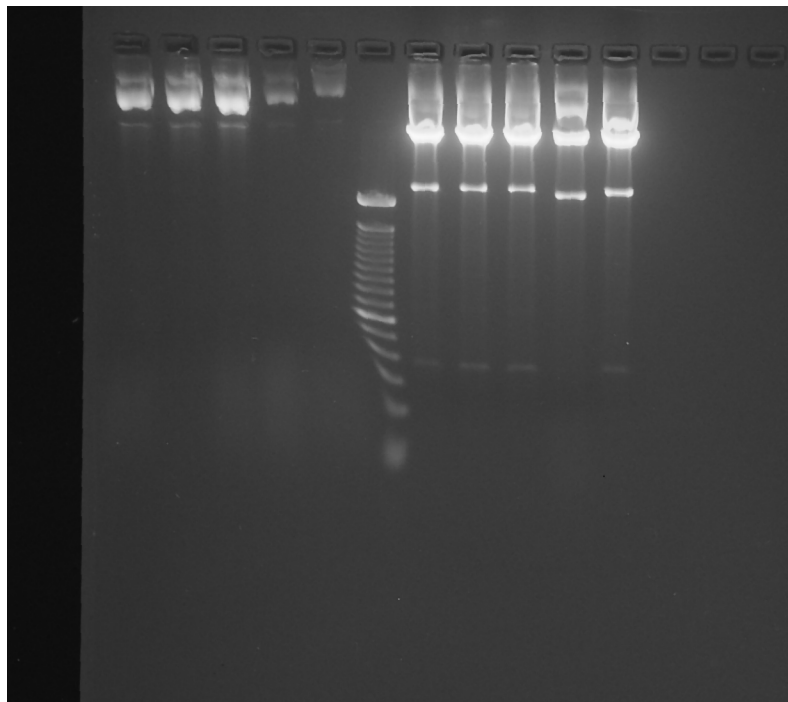


Abbildung 2.4: Gel mit stark verzerrter Lane

3 Lanedetektion

Bevor eine Entzerrung der Lanes stattfinden kann, muss der genaue Verlauf der Lanes bestimmt werden.

3.1 Bildvorverarbeitung

Die Lanedetektion kann durch Faktoren wie geringer Kontrast, Artefakte oder Rauschen erschwert werden. Daher kann eine passende Bildvorverarbeitung die Ergebnisse von Algorithmen zur Lanedetektion deutlich verbessern. (Kum06)

Solche Vorverarbeitungen sollten allerdings an Kopien der Bilder durchgeführt werden, um die Bilddaten nicht für spätere Auswertungsschritte zu verändern.

3.1.1 Rauschunterdrückung

Zur Rauschunterdrückung werden hier Minimumfilter und Maximumfilter verwendet. Dabei wird für jedes Pixel aus einer $M \times M$ Pixel großen Umgebung der kleinste bzw. größte Grauwert ausgewählt und dem Pixel zugewiesen. Die Anwendung des Minimumfilters bezeichnet man als Erosion, die des Maximumfilters wird als Dilatation bezeichnet. (Kum06)

Eine Abfolge von N Erosionen und N Dilatationen wird als Opening bezeichnet. Jede Erosion verkleinert dabei helle Flächen im Bild. Die Dilatationen kehren dies um und vergrößern die hellen Flächen wieder. Kleinere helle Artefakte werden durch die Erosionen vollständig entfernt und können somit von den Dilatationen nicht wiederhergestellt werden. M und N müssen dabei abhängig von der Bildgröße und der Breite der Banden gewählt werden. Bei zu großen Werten können auch Banden aus dem Bild entfernt werden. (NFH07)

Somit entfernt das Opening helle Artefakte und führt generell zu einem homogeneren Hintergrund, während die Banden kaum verändert werden.

3.1.2 Kontrasterhöhung

Die Lanes sind besser detektierbar, je stärker sie sich vom Hintergrund abheben. Daher ist eine Verbesserung des Kontrastes von essentieller Wichtigkeit.

Die Lineare Skalierung kann verwendet werden um den Kontrast zu erhöhen, indem

der gesamte Grauwertbereich ausgeschöpft wird. Dafür müssen zuerst ein minimaler Grauwert g_{min} und ein maximaler Grauwert g_{max} mit $g_{min} < g_{max}$ bestimmt werden. Hierfür können die kleinsten bzw. größten im Bild vorkommenden Grauwerte verwendet werden. Es können aber auch andere Werte verwendet werden, z. B. wenn g_{min} und g_{max} nur Ausreißer sind, die für den eigentlichen Bildinhalt keine Relevanz haben. Dies führt allerdings zu Informationsverlust (NFH07). Die skalierten Grauwerte können wie folgt berechnet werden:

$$f(g) = \left\{ \begin{array}{ll} 0 & \text{für } g < g_{min} \\ 255 & \text{für } g > g_{max} \\ \frac{255}{g_{max} - g_{min}}(g - g_{min}) & \text{für } g_{min} \leq g \leq g_{max} \end{array} \right\}$$

Um zusätzlich zum Erhöhen des Kontrastes noch ein besseres Abheben der Grauwertkanten zu erreichen, kann man eine Hochpassfilterung anwenden. Dabei wird das skalierte Bild $t(x,y)$ mit der zweiten Ableitung des Originalbildes $s''(x,y)$ wie folgt überlagert (NFH07):

$$f(x,y) = t(x,y) - c \cdot s''(x,y)$$

Der Parameter c steuert dabei das Gewicht der Überlagerung.

Die zweite Ableitung des Bildes kann diskret mithilfe des Laplace-Operators ermittelt werden. Dazu können folgende Filterkerne verwendet werden (NFH07):

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} \quad \text{oder} \quad \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

Da die Grauwertkanten verwendet werden um die Lanegrenzen zu finden, ist die Hochpassfilterung von großer Wichtigkeit.

3.2 Detektion von Lanegrenzen

Lanegrenzen zeigen sich im Bild durch vertikale Grauwertkanten. Daher muss im ersten Schritt nach solchen Grauwertkanten gesucht werden.

3.2.1 Finden von Segmenten der Lanegrenzen

Da für das Entzerren der exakte Verlauf der Lanes ermittelt werden muss, ist es nicht sinnvoll über die gesamte Höhe des Bildes nach Grauwertkanten zu suchen. Stattdessen wird das Bild in Streifen gleicher Höhe zerlegt, welche jeweils unabhängig voneinander

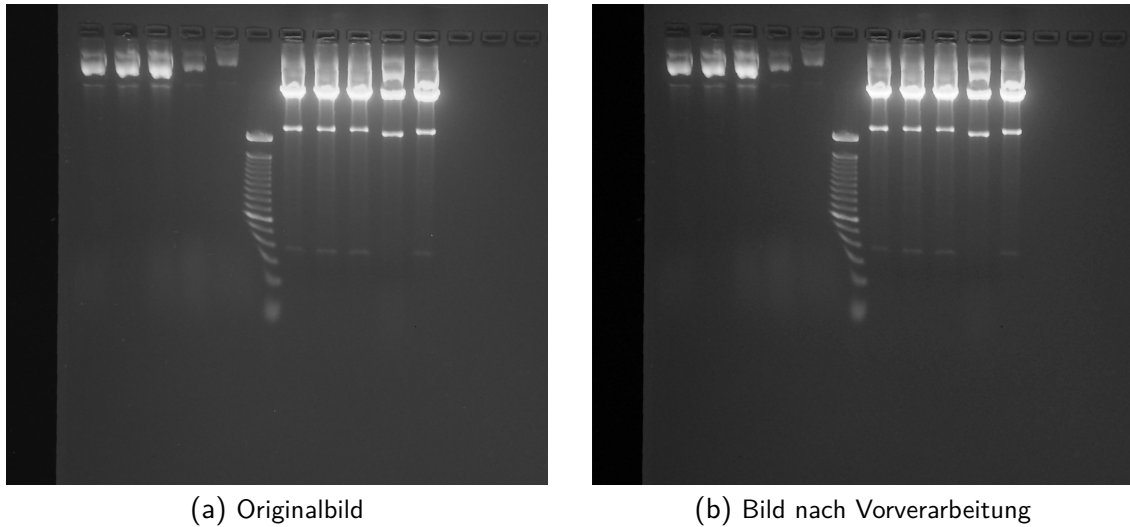


Abbildung 3.1: Bildvorverarbeitung

nach Grauwertkanten durchsucht werden. Die Höhe der Streifen muss dabei so gewählt werden, dass sie klein genug ist um den Laneverlauf erkennen zu können, aber groß genug um kleine Bildstörungen nicht fälschlicherweise als Segmente von Lanegrenzen zu deuten. Außerdem sollte beachtet werden, dass eine hohe Anzahl von Streifen die Verarbeitungszeit recht deutlich ansteigen lässt.

Grauwertkanten können mithilfe von Differenzoperatoren gefunden werden. Da hier nur vertikal verlaufende Grauwertkanten von Interesse sind, kann ein Sobeloperator in y-Richtung verwendet werden (NFH07):

$$H_y = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix}$$

Dadurch zeigen sich die Grenzen der Lanes durch einen hohen Betrag des Grauwertes. Es wird dabei mit vorzeichenbehafteten Grauwerten gearbeitet, da die Vorzeichen später wichtig sind, um linke und rechte Lanegrenzen zu unterscheiden. In jedem Streifen wird unabhängig voneinander nach Segmenten von Lanegrenzen gesucht. Dazu wird für jede Pixelspalte der Mittelwert der Grauwerte der einzelnen Pixel berechnet und darauf ein Schwellwertverfahren angewendet. Dabei sollte beachtet werden, dass zwischen den Banden, sowie am oberen und unteren Bildrand oftmals Streifen existieren, in denen keine Lanegrenzen zu finden sind. Daher muss ein minimaler Schwellwert festgelegt werden, der überschritten werden muss, damit die gefundenen Werte als Segmente von Lanegrenzen anerkannt werden.

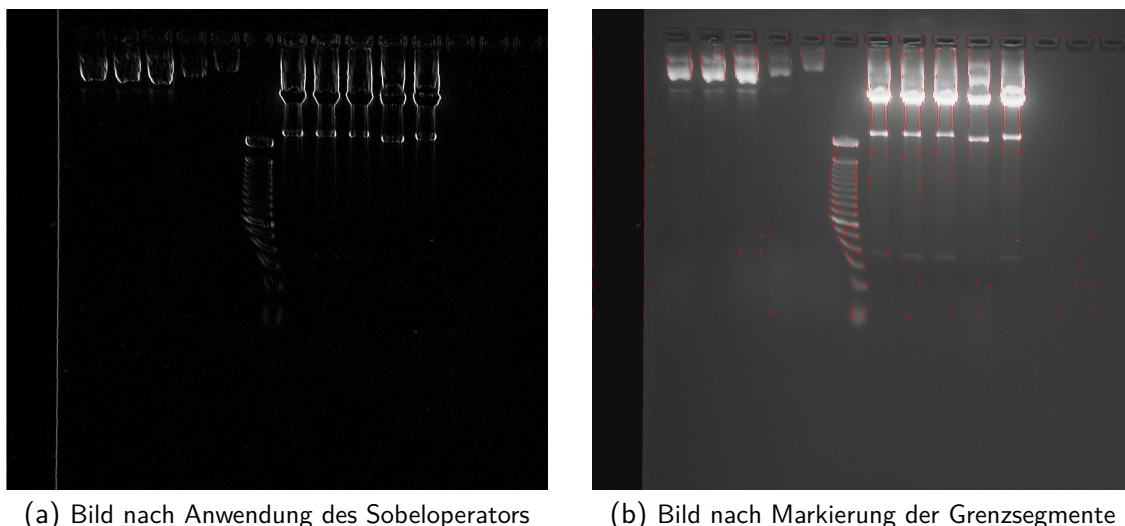


Abbildung 3.2: Segmente der Lanegrenzen

3.2.2 Zusammensetzen der Lanegrenzen

Beim Zusammensetzen der Lanegrenzen werden die einzelnen Streifen von oben nach unten abgearbeitet. In jedem Streifen wird dabei für jedes gefundene Grenzsegment untersucht, ob es zu einer Lanegrenze gehört. Dies ist dann der Fall, wenn es das zur Lanegrenze passende Vorzeichen besitzt und sich innerhalb einer Umgebung von N Pixeln mindestens eines Grenzsegments der Lanegrenze befindet. N muss abhängig von der Bildbreite gewählt werden. Ein zu klein gewähltes N kann dazu führen, dass speziell bei verzerrten Lanes nicht die gesamte Lanegrenze als solche erkannt wird. Ein zu groß gewähltes N kann bei schmalen Lanes zum Zusammenfassen zweier Lanegrenzen zu einer einzelnen führen. Jedes Grenzsegment kann nur einer Lanegrenze zugeordnet werden.

Ein Grenzsegment, welches keiner Lanegrenze zugeordnet werden kann, bildet eine neue Lanegrenze.

3.2.3 Fehlerkorrektur in den Lanegrenzen

Da es aufgrund von Bildstörungen und generellen Inhomogenitäten zwangsläufig zu fälschlicherweise erkannten Grenzsegmenten kommt, und diese auch in die Lanegrenzen aufgenommen werden, findet vor dem Bestimmen der Lanes noch eine Fehlerkorrektur statt. Die folgenden Algorithmen bearbeiten die Grenzsegmente aller Lanegrenzen entsprechend ihrer y -Koordinaten in aufsteigender Reihenfolge.

Eine Lanegrenze kann mehr als ein Grenzsegment in einem Streifen haben, wenn diese Grenzsegmente nahe beieinander liegen. In diesem Fall wird aus den Grenzsegmenten und dem Mittelwert dieser Grenzsegmente jenes ausgesucht, welches dem Laneverlauf

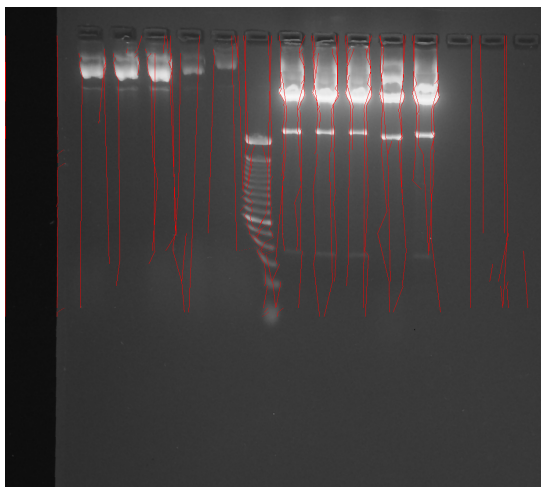
am besten entspricht. Dazu wird mithilfe der Methode der kleinsten Quadrate eine Ausgleichsgerade über die drei vorhergehenden Grenzsegmente berechnet. Es wird also die Gerade $g(y) = x = a \cdot y + b$ gesucht, deren Abstandssumme von den vorhergehenden Grenzsegmenten minimal ist. Es ergibt sich

$$a = \frac{\sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n y_i^2 - n \cdot \bar{y}^2} \quad \text{und} \quad b = \bar{x} - \bar{y} \cdot a$$

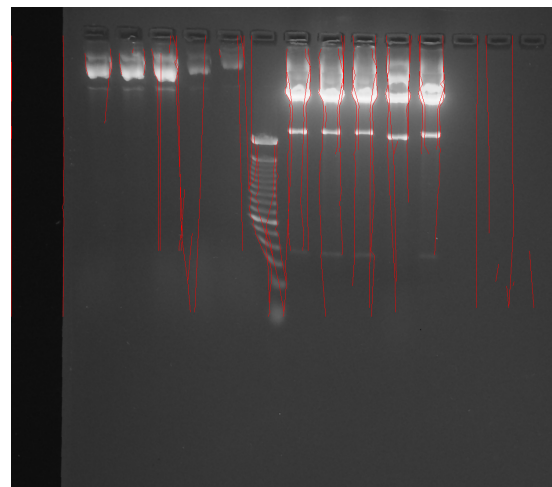
Dabei sind $x_1 \dots x_n$ die x-Koordinaten der Grenzsegmente und $y_1 \dots y_n$ die y-Koordinaten der Grenzsegmente. (Sch12) Nun wird für den Streifen, der bearbeitet wird, die auf der Ausgleichsgeraden liegende x-Koordinate bestimmt. Das Grenzsegment, welches den geringsten Abstand zu dieser optimalen Position hat, wird behalten und alle anderen werden eliminiert.

Als nächstes werden einzelne, vom übrigen Laneverlauf abweichende Grenzsegmente eliminiert. Dabei wird für jedes Grenzsegment ebenfalls wie oben beschrieben eine Ausgleichsgerade mithilfe der Methode der kleinsten Quadrate berechnet. Allerdings werden hier neben den drei vorhergehenden Grenzsegmenten auch die drei folgenden verwendet. Weicht das Grenzsegment zu weit von der Ausgleichsgeraden ab, so wird es eliminiert. Es werden nur Lanegrenzen mit wenigstens vier Grenzsegmenten betrachtet, da sonst nicht sichergestellt werden kann, dass die Ausgleichsgerade wirklich den Laneverlauf wiedergibt.

Zum Schluss werden noch alle Lanegrenzen eliminiert, die nicht ein Minimum an Grenzsegmenten enthalten. Der genaue Wert hängt dabei von der Anzahl der Streifen ab, in die das Bild zerlegt wurde.



(a) Lanegrenzen vor der Fehlerkorrektur



(b) Lanegrenzen nach der Fehlerkorrektur

Abbildung 3.3: Lanegrenzen

3.3 Bestimmung der Lanes

Bevor die Bestimmung der Lanes beginnen kann, werden die Lanegrenzen anhand ihrer x-Koordinaten aufsteigend sortiert.

3.3.1 Finden von Lanekandidaten

Jede Lanegrenze enthält nur Grenzsegmente mit dem gleichen Vorzeichen. Dieses Vorzeichen gibt an, ob es sich um eine linke oder eine rechte Grenze handelt. Das Ziel dieses Schrittes ist es, jeweils eine linke Grenze und eine rechte Grenze mit größerer x-Koordinate zu einer Lane zusammenzufassen. Die Lanes dürfen sich dabei nicht überlappen.

Zu Beginn dieses Schrittes ist keine linke und rechte Grenze ausgewählt. Die Lanegrenzen werden entsprechend ihrer x-Koordinaten nacheinander bearbeitet. Zuerst muss eine linke Grenze ausgewählt werden. Werden mehrere linke Grenzen gefunden, bevor eine rechte Grenze auftritt, so wird die längste von ihnen ausgewählt, d. h. diejenige, welche aus den meisten Grenzsegmenten besteht. Danach muss eine rechte Grenze ausgewählt werden. Auch hier wird die längste ausgewählt, falls mehrere gefunden werden, bevor die nächste linke Grenze auftritt. Die ausgewählte linke und rechte Grenze wird zur einer Lane zusammengefasst. Dieser Algorithmus wird fortgesetzt, bis alle Lanegrenzen bearbeitet wurden.

3.3.2 Prüfung der gefundenen Lanekandidaten auf Richtigkeit

Es ist möglich, dass nicht alle Lanes gefunden wurden oder Lanes gefunden wurden, die eigentlich keine sind. Aus diesem Grund sind weitere Maßnahmen notwendig, um diese Fehler zu beheben.

Wenn sich die Ränder der Gelplatte im Bild befindet, dann kann sie sich als Grauwertkante abzeichnen, die über die gesamte Höhe des Bildes verläuft. Um zu verhindern, dass diese Ränder Lanes bilden, wird der maximale Unterschied zwischen den Längen der beiden Grenzen begrenzt. Des weiteren können an den Rändern der Lanes Unregelmäßigkeiten auftreten, wodurch mehrere dicht nebeneinander liegende Lanegrenzen gefunden werden können. Deshalb müssen alle gefundenen Lanes eine Mindestbreite überschreiten.

Da bei der Gelelektrophorese alle Lanes etwa die gleiche Breite haben, muss dies auch bei den gefundenen Lanes der Fall sein. Also wird der Mittelwert der Breite aller gefundener Lanes ermittelt und alle Lanes eliminiert, die zu weit davon abweichen. Da nun die fehlerhaft erkannten Lanes mit großer Breitenabweichung eliminiert sind, wird erneut der Mittelwert der Breiten der Lanes ermittelt. Bei diesem Durchlauf ist der Toleranzbereich,

in dem die Breite der Lanes liegen muss, um nicht eliminiert zu werden, geringer als beim ersten Durchlauf.

Wenn an den Rändern von Lanes aufgrund von Unregelmäßigkeiten mehrere dicht nebeneinander liegende Lanegrenzen gefunden werden, ist es möglich, dass nicht alle Lanes gefunden werden. Deshalb wird nun mithilfe der zusätzlichen Information über die Breite der Lanes erneut nach Lanes gesucht. Dabei werden nur die Grenzen betrachtet, die noch nicht zu einer Lane gehören. Es wird für jede linke Grenze versucht, eine passende rechte Grenze zu finden, bei der die Breite der entstehenden Lane innerhalb des Toleranzbereichs für die Lanebreite liegt. Können mehrere solcher Lanes gebildet werden, so wird jene ausgewählt, welche die größte gewichtete Länge hat. Die gewichtete Länge L_G ist dabei $L_G = L_L + L_R - \left| \frac{L_L - L_R}{2} \right|$, wobei L_L , L_R die Anzahl der Grenzsegmente innerhalb der linken bzw. rechten Grenze ist. Überschneidet sich eine der neu gefunden Lanes mit einer bereits vorhandenen, so wird diejenige mit der größeren gewichteten Länge behalten und die andere eliminiert.

3.3.3 Suchen der Positionen von nicht erkannten Lanes

Es ist möglich, dass an dieser Stelle noch nicht alle Lanes erfolgreich detektiert wurden. Deshalb werden hier die Positionen möglicher weiterer Lanes ermittelt. Dazu wird zuerst die mittlere Lanebreite \bar{b} berechnet. Danach erfolgt die Berechnung des mittleren Laneabstandes \bar{d} . Dabei wird der Mittelwert aus den Abständen der linken Grenzen benachbarter Lanes berechnet, solange für diese Abstände d gilt: $d \leq 2 \cdot \bar{b}$. Größere Abstände deuten auf dazwischen liegende nicht erkannte Lanes oder Missing Lanes hin. Schließlich wird noch die mittlere Lücke \bar{l} berechnet über $\bar{l} = \bar{d} - \bar{b}$.

Nun wird vor der ersten Lane, nach der letzten Lane und zwischen jeweils zwei benachbarten Lanes nach Lücken l gesucht, für die $l \geq \bar{d}$ gilt. Wird eine solche Lücke gefunden, so wird sie startend bei der rechten Grenze der vorherigen Lane mit einer Breite $w = \bar{b} + 2 \cdot \bar{l}$ in der Liste möglicher Lanepositionen eingetragen. Zwischen zwei benachbarten Lanes können mehrere solche Lücken liegen. Alle bereits detektierten Lanes werden ebenfalls in diese Liste eingetragen.

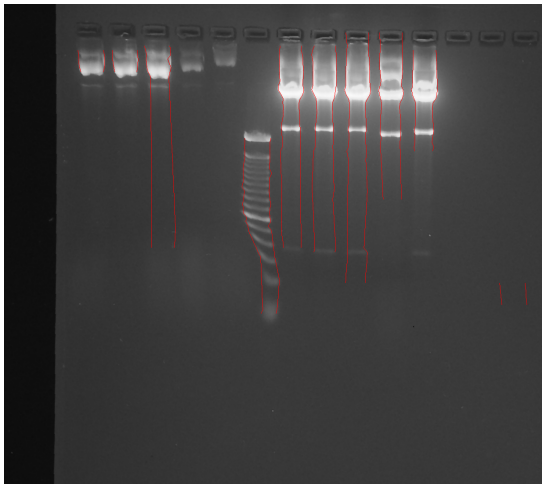
3.3.4 Erneutes Suchen nach Lanes

Bei dem in Kapitel 3.2.1 vorgestellten Verfahren können sehr helle Banden auf einer Lane das Finden von weniger hellen Banden auf der gleichen Höhe in anderen Lanes verhindern. Um dies zu vermeiden, wird dieses Verfahren nun auf jede Lane einzeln angewendet. Die weitere Verarbeitung erfolgt wie in den Kapiteln 3.2.2 und 3.2.3 beschrieben. Da hier die Anzahl der gefunden Lanegrenzen sehr überschaubar ist, reicht es aus, lediglich die jeweils längste linke und rechte Grenzen zu suchen und diese zu einer

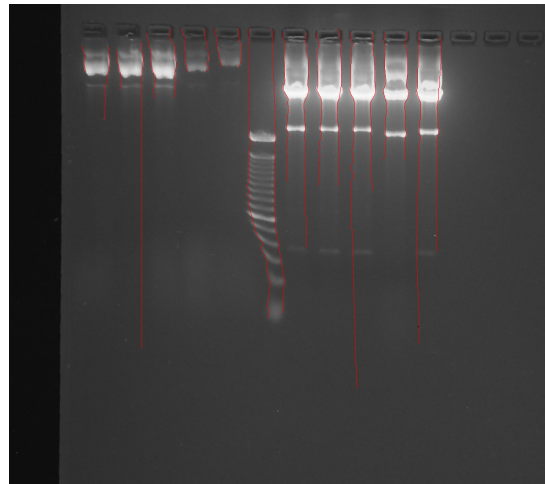
Lane zusammensetzen. Weicht diese Lane nicht zu weit von \bar{b} ab, so wird sie in die Liste der gefunden Lanes aufgenommen.

3.3.5 Abschneiden und Auffüllen der Lanes

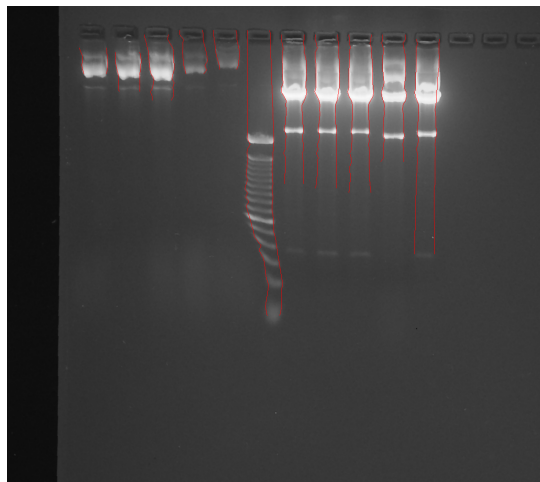
Für die weiteren Bearbeitungsschritte ist es notwendig, dass die beiden Grenzen einer Lane die gleiche Länge haben. Deshalb werden alle Grenzsegmente einer Grenze, die die Länge der anderen Grenze überschreiten, entfernt. Es ist ebenfalls notwendig, dass die Grenzen keine Lücken enthalten. Daher wird nun in allen Streifen, in denen noch kein Grenzsegment vorhanden ist, ein solches hinzugefügt. Der x-Wert des Grenzsegments wird dabei wie in Kapitel 3.2.3 beschrieben mit Hilfe der Methode der kleinsten Quadrate berechnet.



(a) Lanes nach dem ersten Suchlauf



(b) Lanes nach dem zweiten Suchlauf



(c) Lanes nach dem Abschneiden

Abbildung 3.4: Lanes

4 Lane-Entzerrung

4.1 Feststellen der zu entzerrenden Lanes

Zuerst muss festgestellt werden, ob sich überhaupt verzerrte Lanes in dem untersuchten Bild befinden.

Für das Feststellen von Verzerrungen der Lanes wird die Ausrichtung der Lanes verwendet. Für die Berechnung der Ausrichtung wird für jeden Streifen der Mittelwert der x-Werte von linker und rechter Grenze verwendet. Aus diesen Punkten kann wie in Kapitel 3.2.3 beschrieben mit Hilfe der Methode der kleinsten Quadrate eine Gerade der Form $g(y) = x = a \cdot y + b$ berechnet werden. a ist hierbei die Ausrichtung der Lane. Ist $a = 0$, so verläuft die Lane optimal senkrecht. Je größer $|a|$ wird, desto stärker ist die Lane gegenüber dieser optimalen Ausrichtung geneigt.

Danach wird der Abstand zwischen der Gerade g und der linken bzw. rechten Grenze der Lane betrachtet. Eine Verzerrung der Lane liegt dann vor, wenn sich die Abstände der linken und rechten Grenze zu g zu stark unterscheiden. Im Bild zeigt sich dies durch eine Krümmung der Lane. Eine gleichmäßige Änderung beider Abstände kann auf eine zu große Menge des Stoffes und daraus resultierend eine lokale Überladung hinweisen. Dies ist nicht als Verzerrung zu betrachten.

Wird auf diese Weise eine Verzerrung erkannt, so wird diese Lane für die Entzerrung markiert. Es wird außerdem der Beginn der Verzerrung ermittelt und der Bereich von dort bis zum Ende der Lane als zu entzerrender Bereich ausgewählt.

4.2 Bandendetektion

Für das Finden von geeigneten Passpunkten ist es sinnvoll, zuerst die Banden zu detektieren.

4.2.1 Vorbereitung

Zum Finden der Banden werden, genau wie bei der Detektion der Lanes, die Grauwertkanten genutzt. Als Ausgangsbild für diesen Verarbeitungsschritt wird das Originalbild verwendet, das wie in Kapitel 3.1 beschrieben vorverarbeitet wurde. Zum Finden der Grauwertkanten wird hier ebenfalls ein Sobeloperator verwendet. Da für die Bandendetektion allerdings die horizontal verlaufenden Grauwertkanten verwendet werden, wird

ein Sobeloperator in x-Richtung verwendet (NFH07):

$$H_x = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

Es entsteht ein Bild S , dessen vorzeichenbehafteter Grauwert $G_S(x, y)$ die Grauwertänderung in y-Richtung am Punkt (x, y) entspricht. Dabei kennzeichnen negative Werte steigende Grauwertänderungen und positive Werte zeigen fallende Grauwertänderungen.

4.2.2 Finden der Banden

Der Algorithmus bearbeitet den Bereich innerhalb der Lanegrenzen zeilenweise mit aufsteigenden x-Werten und innerhalb jeder Zeile mit aufsteigenden y-Werten. Es wird abwechselnd nach oberen und unteren Kanten gesucht, beginnend mit einer oberen Kante. Wird ein Wert mit dem gesuchten Vorzeichen v gefunden, der den Schwellwert t überschreitet, so ist dieser der erste Punkt einer Kante. Nun wird ausgehend von diesem Punkt nach weiteren Punkten gesucht, die zu dieser Kante gehören. Alle Punkte, für die $|G(x_i, y_i)| \geq t$, sowie $\text{sgn}(G(x_i, y_i)) = v$ und $y_{i-1} - d \leq y \leq y_{i-1} + d$ gilt, werden der Kante hinzugefügt. Dabei sind (x_n, y_n) Punkte, die zur Kante gehören und d der maximale Abstand, den ein Punkt vom vorherigen Punkt haben darf. Wurde die gesamte Breite der Lane nach weiteren, zur Kante gehörenden, Punkten durchsucht, so ist diese Kante fertiggestellt. Überschreitet die Anzahl ihrer Punkte die Mindestanzahl l , so wird nun nach einer Kante des anderen Vorzeichens gesucht, sonst wird sie verworfen und v wird nicht verändert. Danach werden die nächsten d Kanten übersprungen und mit der Suche nach weiteren Kanten fortgefahren. Durch das Überspringen wird verhindert, dass die gleiche Kante mehrfach gefunden wird. Wurde die gesamte Lane auf diese Weise nach Kanten durchsucht, so werden jeweils eine obere und eine untere Kante zu einer Bande zusammengefasst.

Schließlich werden die Kanten aller Banden noch aufgefüllt, es wird also für jeden x-Wert innerhalb der Bande ein Punkt bestimmt, falls diesem noch keiner zugeordnet ist. Dabei wird auf zwei benachbarte Punkte $P_i(x_i, y_i)$ und $P_{i+1}(x_{i+1}, y_{i+1})$ eine Gerade gelegt und für alle Punkte $P_j(x_j, y_j)$ mit $x_i < j < x_{i+1}$ der Punkt ausgewählt, der den geringsten Abstand zur Gerade hat. Dabei gilt:

$$y_j = y_i + \frac{x_j - x_i}{x_{i+1} - x_i} \cdot y_{i+1} - y_i$$

Das Auffüllen ist für die Bestimmung der Passpunkte notwendig.

Die Bandendetektion wird nur für verzerrte Lanes durchgeführt, um die Laufzeit möglichst gering zu halten.

4.3 Lane-Entzerrung

Da die Banden nun detektiert wurden, kann mit der Entzerrung begonnen werden. Zuerst wird allerdings der gesamte zu entzerrende Bereich mit dem Grauwert des Hintergrundes um die Lane überschrieben, da nicht der gesamte zu entzerrende Bereich von der entzerrten Lane verdeckt sein muss und in diesem Fall Reste der verzerrten Lane zurückbleiben können. Diese und auch alle weiteren Operationen am Bild werden an einer Kopie des Originalbildes durchgeführt, um das Resampling aus dem Originalbild nicht zu stören.

4.3.1 Verlauf der Entzerrten Lane

Als nächstes muss der Verlauf der entzerrten Lane bestimmt werden. Dazu wird die Ausrichtung des unverzerrten Teils der Lane berechnet, wie in Kapitel 4.1 beschrieben. Ist der unverzerrte Teil der Lane nicht lang genug um zuverlässig eine Ausrichtung berechnen zu können oder ist die gesamte Lane verzerrt, so wird der Mittelwert der Ausrichtungen aller nicht verzerrter Lanes stattdessen verwendet. Mit Hilfe der Geradengleichung $g(y) = x = a \cdot y + b$, wobei a die Ausrichtung ist, kann nun der weitere Verlauf beider Lanegrenzen ermittelt werden.

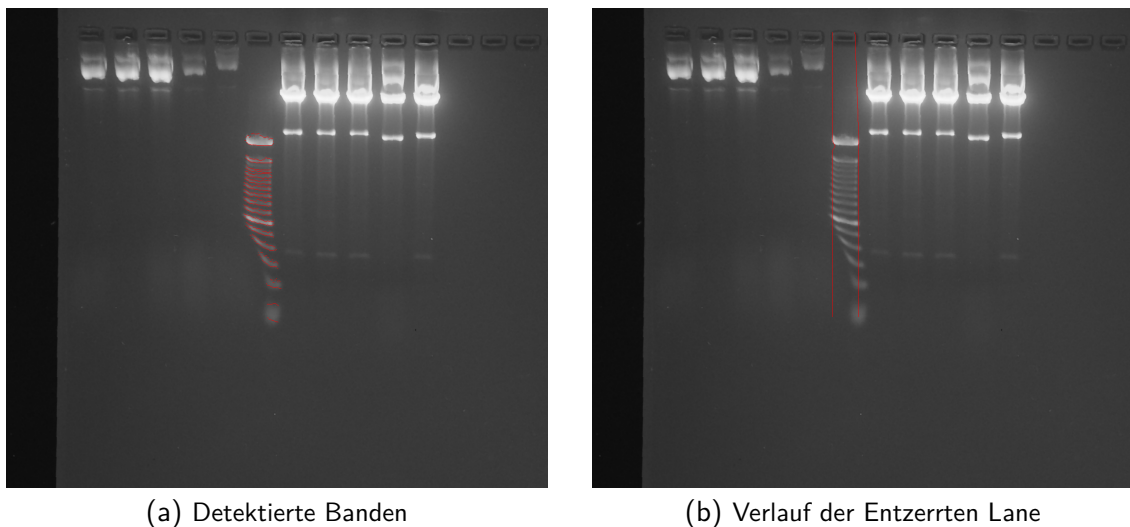


Abbildung 4.1: Vorbereitung der Entzerrung

4.3.2 Auswahl der Passpunkte

Für die Entzerrung der Lane müssen zuerst geeignete Passpunkte ausgewählt werden. Diese Passpunkte werden für jede Bande einzeln bestimmt. Zuerst werden die Punkte in der verzerrten Lane gesucht. Die ersten vier Passpunkte sind die vier Ecken der Bande, d. h. jeweils die Punkte mit dem kleinsten und dem größten x-Wert in der oberen und der

unteren Kante. Zusätzlich werden noch die Mittelwerte der oberen und unteren linken Ecke, sowie der Mittelwert der oberen und unteren rechten Ecke ausgewählt. Schließlich werden auf den Kanten noch jeweils vier weitere Passpunkte ausgewählt. Die x-Werte werden dabei wie folgt berechnet:

$$x_i = (i + 1) \cdot \frac{x_{\max} - x_{\min}}{5} + x_{\min} \text{ für } 0 \leq i < 4$$

Dabei sind x_{\min} und x_{\max} der kleinste bzw. größte in der Kante vorkommende x-Wert. Es werden somit insgesamt 14 Passpunkte ausgewählt.

Nun werden die entsprechenden Punkte in der entzerrten Lane gesucht. Es wird angenommen, dass die Banden horizontal verlaufen sollen. Weiterhin wird angenommen, dass das Verhältnis des Abstandes eines Punktes vom linken Rand der Kante zur Länge der Kante konstant bleibt.

$$\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} = \frac{x'_i - x'_{\min}}{x'_{\max} - x'_{\min}}$$

wobei x'_i , x'_{\min} und x'_{\max} der x-Wert des Passpunktes, der kleinste x-Wert bzw. der größte x-Wert in der entzerrten Lane sind. Daraus ergibt sich:

$$x'_i = \frac{(x_i - x_{\min}) \cdot (x'_{\max} - x'_{\min})}{x_{\max} - x_{\min}} + x'_{\min}$$

Weiterhin wird angenommen, dass zumindest ein Teil der Bande nicht verzerrt ist. Deshalb werden als y-Werte die Modalwerte der jeweiligen Bande verwendet.

4.3.3 Zerlegung der Lane

Die Entzerrung soll in verschiedenen Abschnitten der Lane getrennt voneinander durchgeführt werden, wobei jeder Abschnitt nur Banden mit ähnlicher Ausrichtung enthalten soll. Dazu muss zuerst die Ausrichtung der Banden ermittelt werden. Dabei wird für beide Kanten aus allen allen Punkten der Kanten mit der Methode der kleinsten Quadrate eine Ausgleichsgerade $g(x) = y = a \cdot x + b$ gesucht. Es ergibt sich (Sch12):

$$a = \frac{\sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2} \quad \text{und} \quad b = \bar{y} - \bar{x} \cdot a$$

Dabei sind $x_1..x_n$ und $y_1..y_n$ die x- bzw. y-Koordinaten der einzelnen Punkte. Der Anstieg a der Geraden ist die Ausrichtung der Kante. Die Ausrichtung der Bande ist dabei der Mittelwert aus den Ausrichtungen der Kanten.

Nun werden benachbarte Banden mit ähnlichen Ausrichtungen zu Abschnitten zusam-

mengefasst, die gemeinsam transformiert werden.

4.3.4 Transformation

Zur Entzerrung der Lane wird eine stückweise Polynomtransformation mit Polynomen des ersten Grades verwendet. Allgemein kann die Polynomtransformation folgendermaßen definiert werden (Can02):

$$X = \sum_{i=0}^n \sum_{j=0}^{n-1} C_{ij} \cdot x^i \cdot y^j$$

$$Y = \sum_{i=0}^n \sum_{j=0}^{n-1} C'_{ij} \cdot x^i \cdot y^j$$

Dabei sind x, y die ursprünglichen Koordinaten, X, Y die transformierten Koordinaten, n der Grad des verwendeten Polynoms und C_{ij} und C'_{ij} die Koeffizienten, die zu definieren sind, um eine günstige Transformation zu erhalten. Daraus ergibt sich für Polynome des ersten Grades:

$$X = C_{00} + C_{01} \cdot y + C_{10} \cdot x$$

$$Y = C'_{00} + C'_{01} \cdot y + C'_{10} \cdot x$$

Es ergibt sich also für die Transformation der x- und y-Koordinaten jeweils ein überbestimmtes Gleichungssystem $A \cdot x = b$. A enthält dabei die Werte von x und y , b enthält die Werte von X und Y und x enthält die gesuchten Werte von C_{ij} und C'_{ij} . Zur näherungsweisen Bestimmung von x wird die Methode der kleinsten Quadrate mit QR Zerlegung angewendet (Für Details siehe (GVL13)).

4.3.5 Resampling

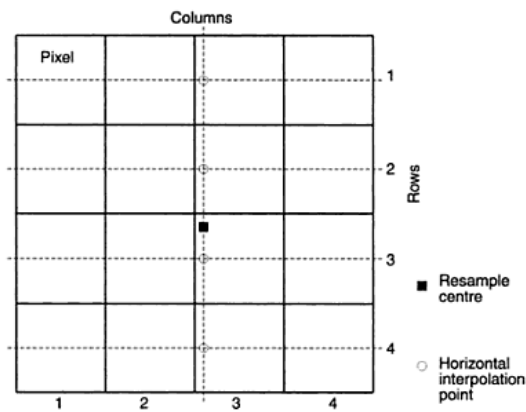
Da nun die Koeffizienten C_{ij} und C'_{ij} berechnet wurden, kann die Transformation begonnen werden. Es wird hier eine Rückwärtstransformation durchgeführt, d. h. für jedes Pixel der entzerrten Lane wird das entsprechende Pixel im unkorrigierten Bild berechnet (Jen05).

Für die Ermittlung des Grauwerts wird die Kubische Faltung (Cubic Convolution) verwendet. Dabei wird ein Bereich von 4x4 Pixel um die berechnete Koordinate in die Rechnung einbezogen. Zuerst werden die Mittelwerte der einzelnen Zeilen berechnet, im zweiten Schritt erfolgt dann die Berechnung des endgültigen Grauwerts aus den vier

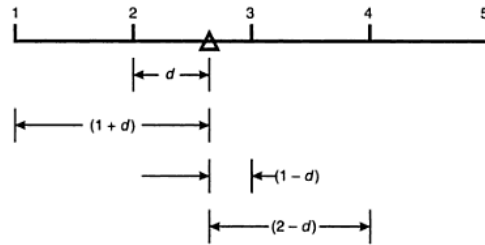
Mittelwerten der Zeilen. Die Mittelwerte werden dabei wie folgt berechnet (McC05):

$$\begin{aligned}
 R_i = & R_1 \cdot (4 - 8 \cdot (1 + d) + 5 \cdot (1 + d)^2 - (1 + d)^3) + \\
 & R_2 \cdot (1 - 2 \cdot d^2 + d^3) + \\
 & R_3 \cdot (1 - 2 \cdot (1 - d)^2 + (1 - d)^3) + \\
 & R_4 \cdot (4 - 8 \cdot (2 - d) + 5 \cdot (2 - d)^2 - (2 - d)^3)
 \end{aligned}$$

R_1, R_2, \dots sind nummeriert wie in Abbildung 4.2a beschrieben. d ist der Abstand zu Pixelmittelpunkten und ist in Abbildung 4.2b beschrieben.



(a) Interpolationsachsen (McC05)



(b) Abstände zu Pixelmittelpunkten (McC05)

Abbildung 4.2: Kubische Faltung

Nach Abschluss der Transformation entsteht das entzerrte Bild.

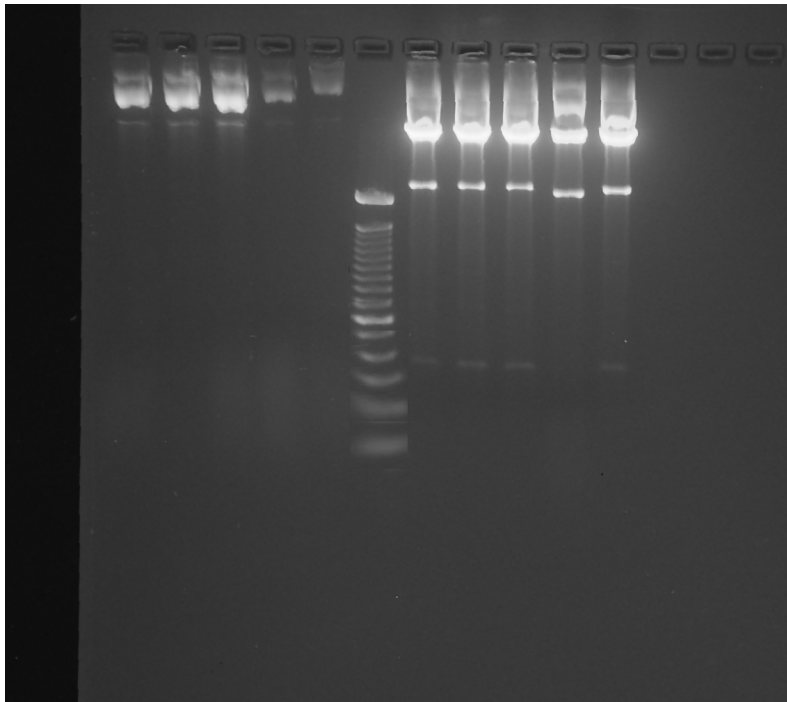


Abbildung 4.3: Entzerrtes Bild

5 Überblick über die Implementierung

5.1 Aufruf

Beim Aufruf müssen dem Algorithmus die Bilddaten, die Höhe und Breite sowie die Farbtiefe des Bildes übergeben werden. Zusätzlich können noch verschiedene Einstellungen übergeben werden, mit denen die Verarbeitung des Algorithmus beeinflusst werden kann, z. B. die Filtergröße für die Bildvorverarbeitung oder die Höhe der einzelnen Streifen, in die das Bild für die Lanedetektion zerlegt wird.

Als Ergebnis würde der Algorithmus dann die Information, ob eine Transformation stattgefunden hat, sowie das transformierte Bild zurückgeben.

5.2 Übersicht über die Klassen

Die Band-Klasse repräsentiert eine einzelne Bande. Es sind die obere und untere Kante als Liste von Punkten, sowie verschiedene Metriken zur Beschreibung der Bande enthalten. Alle Operationen, die eine einzelne Bande betreffen, z. B. die Berechnung der Ausrichtung oder das Finden von Kontrollpunkten, sind hier implementiert.

Die LaneBoundary-Klasse repräsentiert eine einzelne Lanegrenze. Es sind neben den einzelnen enthaltenen Grenzsegmenten als Liste von Punkten auch die Information, ob es sich um eine linke oder rechte Bande handelt, sowie verschiedene Metriken zur Beschreibung der Lanegrenze enthalten. Alle Operationen, die eine einzelne Lanegrenze betreffen, z. B. die in Kapitel 3.2.3 beschriebenen Fehlerkorrekturen oder die Berechnung des neuen Verlaufs in der entzerrten Lane, sind hier implementiert.

Die Lane-Klasse repräsentiert eine einzelne Lane. Hier sind die beiden Lanegrenzen sowie verschiedene Metriken zur Beschreibung der Lane enthalten. Wurde eine Bandendetektion durchgeführt, so sind auch alle Banden enthalten. Alle Operationen, die eine einzelne Bande betreffen, z. B. die Bandendetektion oder die Entzerrung der Lane, sind hier implementiert. Teile dieser Operationen werden von den enthaltenen Band- und LaneBoundary-Objekten ausgeführt.

Die GrayscaleImage-Klasse repräsentiert ein einzelnes Graustufenbild. Die Bilddaten sowie Höhe, Breite und Farbtiefe des Bildes sind hier enthalten. Es werden mehrere dieser Objekte benötigt, da einige Operationen das Bild auf eine Weise verändern, die es für nachfolgende Operationen unbrauchbar macht. Alle Operationen, die direkt auf das Bild ausgeführt werden, z. B. das Anwenden von Differenzoperatoren, das Suchen der Grenzsegmente oder das Resampling, sind hier implementiert.

Die LaneRectificator-Klasse ist die übergeordnete Steuerung des Algorithmus. Hier sind alle Bilder und Lanes enthalten und der gesamte Ablauf des Algorithmus wird hier gesteuert.

6 Auswertung

6.1 Zusammenfassung

Die vorliegende Bachelorarbeit befasst sich mit Entwicklung und Implementierung eines Algorithmus zur automatisierten Entzerrung von Gelelektrophorese-Blots. Dabei wurden zuerst die physikalischen Grundlagen, sowie Auswertungsverfahren und Anwendungsgebiete der Gelelektrophorese erläutert. Danach wurde ein Verfahren zur Lanedetektion vorgestellt. Da für das Erkennen von Verzerrungen nicht nur die Position der Lanes bekannt sein muss, sondern der exakte Verlauf, wurde hier ein Verfahren entwickelt, dass die Lanes über ihre gesamte Länge untersucht. Zum Finden der Lanegrenzen wurden die Grawertkanten genutzt, die mit Hilfe eines Differenzoperators sichtbar gemacht wurden. Anschließend wurden die detektierten Lanes auf Verzerrungen untersucht. Wurden Verzerrungen festgestellt, so wurde noch eine Bandendetektion durchgeführt und der neue, entzerrte Verlauf der Lane wurde bestimmt. Die Entzerrung erfolgte dann durch stückweise Anwendung einer Polynomtransformation mit Polynomen ersten Grades. Durch die Verwendung einer Rückwärtstransformation war die Auswahl einer Resamplingvorschrift notwendig. Hier wurde die Kubische Faltung gewählt, da diese eine gute Qualität liefert und der höhere Rechenaufwand nicht so schwerwiegend ist, da nur ein relativ kleiner Ausschnitt des Bildes transformiert wird.

6.2 Ausblick

Um ein größeres Spektrum an Bildern analysieren zu können, ist es möglich, den Algorithmus um die Verarbeitung von Farbbildern zu erweitern. Dazu müssen die Farbbilder zuerst zu Graustufenbildern konvertiert werden. Die Grauwerte können beispielsweise über die Mittelwerte der einzelnen Farbkanäle abgeschätzt werden. Die weitere Verarbeitung kann dann wie beschrieben ablaufen. Lediglich beim Resampling müssten die Farbbilder auf andere Weise bearbeitet werden. Es müssten dabei die einzelnen Farbkanäle getrennt voneinander durch den Resampler bearbeitet werden.

Literaturverzeichnis

- [Ana98] ANALYSIS, Technical University of D. s.: *The Molecular Structure and Replication of the Genetic Material or Commonly Used METHODS for the Characterisation of DNA Sequences*. http://genome.cbs.dtu.dk/staff/dave/roanoke/fig5_33.jpg, 1998. – abgerufen am 20. 08. 2014
- [Bio09] BIOLOGY, West Chester U. o.: *The Behavioral and Toxic Effects of Sarco-phine in the Planarian Flatworm*. <http://darwin.wcupa.edu/beneski/bio-515/f09/mohanty/Main/ThesisProposal>, 2009. – abgerufen am 20. 08. 2014
- [Can02] CANTERS, F.: *Small-Scale Map Projection Design*. Taylor & Francis, 2002 (Research monographs in geographic information systems). – ISBN 9780203472095
- [GVL13] GOLUB, G.H. ; VAN LOAN, C.F.: *Matrix Computations*. Johns Hopkins University Press, 2013 (Johns Hopkins Studies in the Mathematical Sciences). – ISBN 9781421407944
- [HH11] HAMES, D. ; HOOPER, N.: *BIOS Instant Notes in Biochemistry*. Taylor & Francis, 2011. – ISBN 9781136667879
- [HM J] HEMSCHMEIER, K. S. ; MAELICKE, A.: *Gelelektrophorese*. <http://www.chemgapedia.de/vsengine/vlu/vsc/de/ch/8/bc/vlu/proteinanalytik/elektrophorese.vlu.html>, o. J.. – abgerufen am 20. 08. 2014
- [Jen05] JENSEN, John R.: *Introductory Digital Image Processing: A Remote Sensing Perspective*. Prentice Hall, 2005 (Prentice Hall series in geographic information science). – ISBN 9780131453616
- [KG05] KUMAR, A. ; GARG, N.: *Genetic Engineering*. Nova Biomedical Books, 2005 (Nova biomedical books). – ISBN 9781594547539
- [Kni06] KNIPPERS, R.: *Molekulare Genetik*. Thieme, 2006. – ISBN 9783134770094
- [Kum06] KUMMICH, J.: *Entwicklung von Algorithmen zur Auswertung von eindimensionalen Elektrophoresegelelen*, Westsächsische Hochschule Zwickau (FH), Diplomarbeit, 2006
- [LTD⁺13] LEE, M. ; TIETZE, K.J. ; DUPONT, P.F. ; CHYKA, P.A. ; MACKICHAN,

- J.J. ; MCGORY, R. ; LAU, A. ; CHAN, L.N. ; BEITELSHEES, A.L. ; ORBACH, R.C. u. a.: *Basic Skills in Interpreting Laboratory Data, 5th edition*. ASHP, 2013. – ISBN 9781585283453
- [McC05] MCCLOY, K.R.: *Resource Management Information Systems: Remote Sensing, GIS and Modelling, Second Edition*. Taylor & Francis, 2005. – ISBN 9780415263405
- [NFH07] NISCHWITZ, A. ; FISCHER, M. ; HABERÄCKER, P.: *Computergrafik und Bildverarbeitung*. Vieweg, 2007
- [Ric03] RICHTER, G.: *Praktische Biochemie: Grundlagen und Techniken*. Thieme, 2003. – ISBN 9783131323811
- [Sch10] SCHIERJOTT, K.P.: *Miniaturisierte Kapillarelektrophorese zur kontinuierlichen Überwachung von Kationen und Anionen in Prozessströmen*. KIT Scientific Publ., 2010 (Schriften des Instituts für Mikrostrukturtechnik am Karlsruher Institut für Technologie). – ISBN 9783866445239
- [Sch12] SCHUBERT, M.: *Mathematik Für Informatiker: Ausführlich Erklärt Mit Vielen Programmbeispielen und Aufgaben*. Vieweg+Teubner Verlag, 2012. – ISBN 9783834819956
- [Wes06] WESTERMEIER, R.: *Electrophoresis in Practice*. Wiley, 2006. – ISBN 9783527604449
- [WKC90] WARR, G.W. ; KARAM, J.D. ; CHAO, L.: *Methods in Nucleic Acids Research*. Taylor & Francis, 1990. – ISBN 9780849353116

Erklärung

Hiermit erkläre ich, dass ich meine Arbeit selbstständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die Arbeit noch nicht anderweitig für Prüfungszwecke vorgelegt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Mittweida, 25. August 2014